

# NOTAS

- By Year:
  - Avg\_Times\_Cited:  $\text{Times\_Cited} / \text{Num\_Documents}$
- By Term (Terms analysis):
  - Avg\_Times\_Cited:  $\text{Times\_Cited} / \text{Numero\_de\_años\_desde\_la\_primera\_aparición}$
  - H\_index
  - M\_index y G\_index: [https://en.wikipedia.org/wiki/Author-level\\_metrics#List\\_of\\_metrics](https://en.wikipedia.org/wiki/Author-level_metrics#List_of_metrics)
- SD: Single publication. Publicación de un único autor (author, institution, country)
- MD: Multiple publication. Publicación de varios autores
- SMR: MD / SD. Medida de colaboración entre autores, instituciones o países

Num Authors	%	Acum Num Authors	% Acum	Documents written per Author	Num Documents	% Num Documents	Acum Num Documents	% Acum Num Documents
2	0.11 %	2	0.11 %	6	12	1.76%	12	1.76%
6	0.32 %	8	0.43 %	5	25	3.68%	37	5.44%
15	0.81 %	23	1.24 %	4	35	5.15%	72	10.59%
31	1.67 %	54	2.91 %	3	23	3.38%	95	13.97%
143	7.7 %	197	10.61 %	2	125	18.38%	220	32.35%
1660	89.39 %	1857	100.0 %	1	460	67.65%	680	100.0%

Dos autores han escrito seis documentos cada uno

(CORE AUTHORS) 11% de los autores han escrito el 18% de los artículos

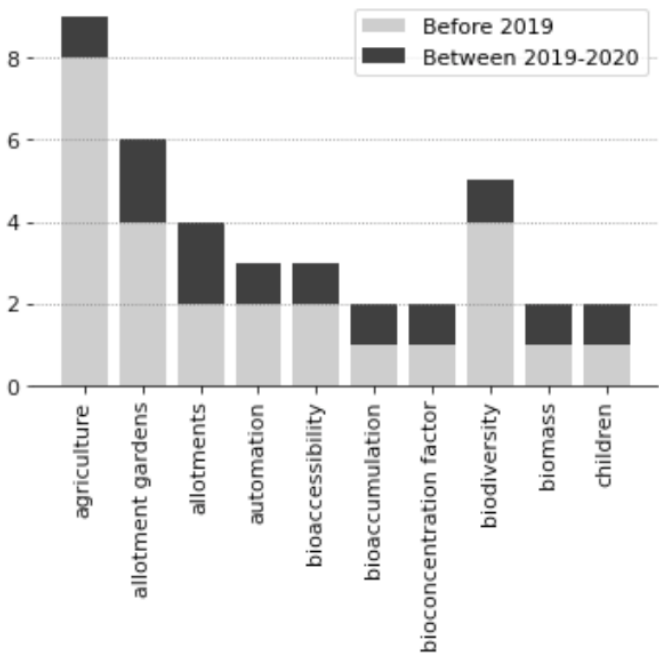
	Num Sources		%	Acum	Num Sources		%	Acum	Documents published		Tot Documents published	Num Documents		Tot Documents	Bradford's Group
0	1		0.24 %		1		0.24 %		29		29	29		4.22 %	1
1	1		0.24 %		2		0.48 %		26		26	55		7.99 %	1
2	1		0.24 %		3		0.72 %		20		20	75		10.9 %	1
3	1		0.24 %		4		0.95 %		15		15	90		13.08 %	1
4	1		0.24 %		5		1.19 %		11		11	101		14.68 %	1
5	1		0.24 %		6		1.43 %		10		10	111		16.13 %	1
6	2		0.48 %		8		1.91 %		9		18	129		18.75 %	1
7	4		0.95 %		12		2.86 %		8		32	161		23.4 %	1
8	1		0.24 %		13		3.1 %		7		7	168		24.42 %	1
9	2		0.48 %		15		3.58 %		6		12	180		26.16 %	1
10	3		0.72 %		18		4.3 %		5		15	195		28.34 %	1
11	6		1.43 %		24		5.73 %		4		24	219		31.83 %	1
12	14		3.34 %		38		9.07 %		3		42	261		37.94 %	2
13	46		10.98 %		84		20.05 %		2		92	353		51.31 %	2
14	335		79.95 %		419		100.0 %		1		335	688		100.0 %	3

Dos revistas han publicado 9 artículos cada una

CORE SOURCES: 20% de las revistas han publicado el 51% de los artículos

Los artículos se dividen en tres partes iguales y se forman tres grupos.  
El primer grupo es el core de las fuentes

	AGR	ADY	PDLY	Before 2019	Between 2019-2020
Author_Keywords					
agriculture	-1.0	0.5	0.07	8	1
allotment gardens	-1.0	1.0	0.15	4	2
allotments	-1.0	1.0	0.15	2	2
automation	-0.5	0.5	0.07	2	1
bioaccessibility	0.0	0.5	0.07	2	1
bioaccumulation	0.5	0.5	0.07	1	1
bioconcentration factor	0.5	0.5	0.07	1	1
biodiversity	0.5	0.5	0.07	4	1
biomass	0.0	0.5	0.07	1	1
children	0.5	0.5	0.07	1	1



AGR: Average growth rate

ADY: Average documents per year

PDLY: Percentage of documents in last years

$$AGR = \frac{\sum_{i=Y_s}^{Y_e} P_i - P_{i-1}}{(Y_e - Y_s) + 1}$$

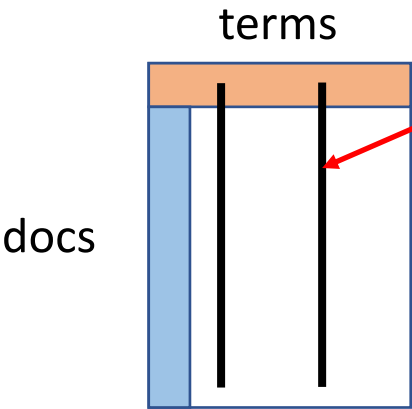
$$ADY = \frac{\sum_{i=Y_s}^{Y_e} P_i}{(Y_e - Y_s) + 1}$$

$$PDLY = \frac{\sum_{i=Y_s}^{Y_e} P_i}{(Y_e - Y_s + 1) * TND} * 100\%$$

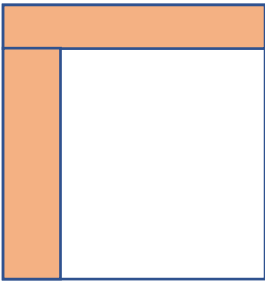
TDN: numero total de documentos

Ventana temporal

# Correlación

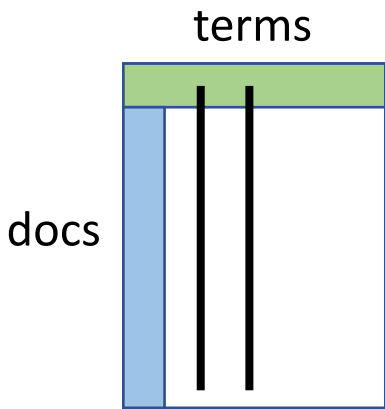


Autocorrelacion: correlación entre las columnas

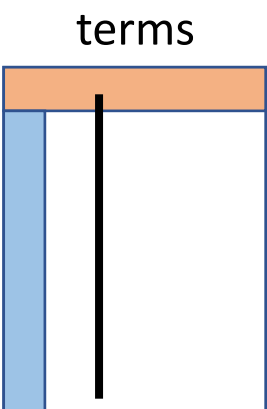


Relación entre los términos de una columna

TF\_matrix: term-frequency



A

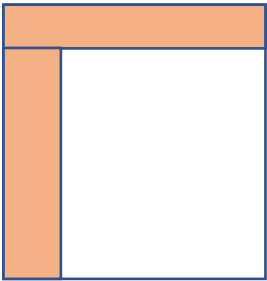
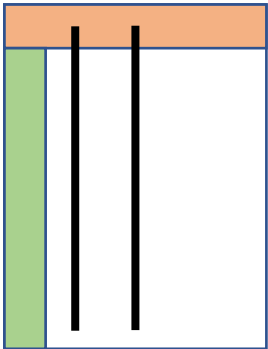


B

Co-ocurrencia

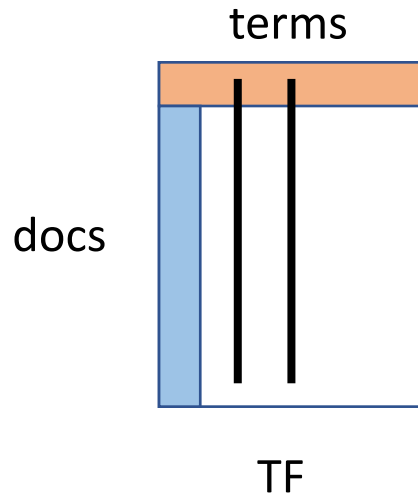
$$A^T \times B$$

Cross-correlacion: correlación entre las columnas



# Matriz TF\*IDF

[https://scikit-learn.org/stable/modules/generated/sklearn.feature\\_extraction.text.TfidfTransformer.html](https://scikit-learn.org/stable/modules/generated/sklearn.feature_extraction.text.TfidfTransformer.html)



Sublinear\_tf == False:

$$\text{tf-idf}(t, d) = \text{tf}(t, d) * \text{idf}(t)$$

Sublinea\_tf == True:

$$\text{tf-idf}(t, d) = [1 + \log(\text{tf}(t, d))] * \text{idf}(t)$$

smooth\_idf == False:

$$\text{idf}(t) = \log [ n / \text{df}(t) ] + 1$$

smooth\_idf == True:

$$\log [ (1 + n) / (1 + \text{df}(t)) ] + 1$$

$\text{df}(t)$  : total de documentos en que aparece t

norm: normaliza cada fila de la matriz resultante

# Selección de keywords

1. Selecciona los términos con una ocurrencia mínima (1, 2, ....)
2. Computa la matriz TF\*IDF y el peso total de cada término: la suma de las columnas de TF\*IDF equivale a un conteo de frecuencia
3. Selecciona los N términos más importantes

## Análisis de Factores

1. Matriz TF
2. Transpuesta
3. PCA
4. Representación gráfica

## SVD de la matriz TF\*IDF

1. Matriz TF
2. Matriz TF\*IDF
3. Transpuesta
4. Aplica SVD a las columns
5. Representacion gráfica

## Análisis Comparativo

1. Matriz TF
2. Análisis de Correspondencias
3. Clustering
4. Representación Gráfica

## Análisis Temático

1. Matriz TF
2. Matriz TF\*IDF
3. Clustering
4. Contingency table
5. Transpuesta
6. Representación gráfica usando CA

## Mapeo de conceptos

1. Co-ocurrencia
2. Normalizacion
3. Clustering herarquico
4. Matriz cluster x cluster
5. Representación gráfica usando MDS

